5Ф

УДК 519.682.5

П.І. Федорук
Прикарпатський національний університет ім. Василя Стефаника,
м. Івано-Франківськ, Україна
pavlo@pu.if.ua

Використання адаптивних тестів в інтелектуальних системах контролю знань

У статті .розглянуто проблеми тестового контролю знань у сучасних навчальних системах. Методики та інструментарій класичної теорії тестів попри великі можливості даної технології не вирішують багатьох задач, які ставить сучасний рівень розвитку освіти. Адаптивні тести у даному випадку дозволяють вирішувати ці питання і можуть бути ефективно використані для вирішення будь-яких задач оптимізації навчального процесу – оцінки ефективності педагогічних інновацій і технологій, моніторингу і т.д. Розглянуто підхід до конструювання тестів, представлений в сучасній теорії тестування, на основі математичної теорії параметричної оцінки тестових завдань на базі основних сучасних математичних моделей: однопараметричної моделі Раша, двопараметричної та трипараметричної моделі Бірнбаума.

Вступ

Адаптивне тестування – різновид тестування, при якому порядок представлення запитань (або складність) залежить від відповідей того, хто тестується, на попередні запитання [1], [2].

Використання адаптивних тестів дозволяє:

– підлаштовуватися під індивідуальні можливості студента – виключаються завдання, які або занадто складні, або занадто легкі;

– підвищити точність оцінки рівня знань сильних і слабких студентів завдяки використанню більшого банку запитань різного рівня складності;

– зменшити тривалість тесту і кількість поставлених запитань, необхідних для досягнення достатньої точності оцінки рівня знань студента;

– знизити ступінь втомленості студента;

– забезпечити конфіденційність за рахунок надання кожному студентові індивідуального набору тестових завдань, що відповідають його рівню знань;

– спростити процедуру внесення змін у банк тестових завдань, які будуть автоматично враховані адаптивним алгоритмом.

1. Адаптивне тестування і класична теорія тестів

Адаптивне тестування може бути організовано по-різному. Найпростіший варіант – це приписати кожному тестовому запитанню міру складності. При правильній відповіді студента наступне запитання буде більш складним, при неправильній – менш складним. Також тестові запитання можуть бути прив’язані до певних тем, навчальних матеріалів. Якщо у ході відповіді студент зробить помилку, наступне запитання буде з тієї теми, яку він не досить добре знає. Якщо студент правильно відповість на запитання, йому пропонується запитання з іншої теми. При адаптивному тестуванні

може використовуватись тимчасова модель студента, в яку заносяться дані про успішність студента в поточний момент. Після закінчення тесту можливе оновлення загальної моделі.

Перехід від традиційної класичної теорії тестів до адаптивного тестування в практичному плані неминуче пов'язаний з деякими труднощами:

– з необхідністю залучення досить складного математичного апарату;

– з використанням дорогої комп'ютерної техніки і розробки спеціальних програмних продуктів;

– з необхідністю проведення ретельної роботи зі стратифікації вибірки випробовуваних статистик з математичними моделями вимірювання.

Порівняємо класичну теорію тестування і теорію адаптивного тестування (IRT) (табл. 1):

Таблиця 1 – Порівняння класичної теорії тестування і IRT

Класична теорія	Item Response Theory
Основне завдання тесту: Дізнатися значення дійсного бала (Т) учасника тестування, виходячи із результату (Х) та з врахуванням помилки вимірювання (Е). Отримаємо основний постулат класичної теорії тестування: Xi = Ti + Ei.	Основне завдання тесту: Отримати стійку об'єктивну оцінку латентного параметра рівня знань учасника тестування з досліджуваного предмета.
Можливості:	
Отримати інформацію про тест, на підставі матриці результатів апробаційного тестування, а саме: 1. Оцінити статистичну складність завдань.	1. Встановлення зв'язку між патентними параметрами учасника тестування і спостережуваними результатами виконання тесту. Розглядається взаємодія двох латентних параметрів тесту: рівня знань учасника тестування і складності завдань.
2. Інтеркореляція між завданнями тесту і кореляцією балів завдань і зовнішнього критерію (суми балів випробовуваних) для визначення валідності тестових завдань.	2. Параметри рівня знань учасника тестування і складності завдань тесту відображаються в єдину шкалу логітів, що дозволяє реалізувати ідею адаптивного тестування, коли для кожного випробовуваного (з конкретним рівнем знань) вибирають завдання певної складності.
3. Швидко оцінити якість тесту на основі графічного виду кривої розподілу тестових балів учасників тестування.	3. Вводиться новий критерій ефективності тесту – інформаційна функція.
4. Отримаємо оцінку надійності результатів тестування за допомогою кореляційного аналізу балів учасників тестування.	4. У Item Response Theory реалізована можливість застосування не тільки рейтингової, але й інтервальної шкали, а це означає: рівень знань учасника тестування можна оцінити кількісно.
5. Побудувати довірчий інтервал, в межах якого знаходиться дійсний бал випробовуваного, або отримати точкову регресійну оцінку.	

Цілі адаптивного тестування:

– орієнтація в навчальному процесі на адаптивні методи навчання і контролю;

– організація на самостійну роботу учнів за допомогою адаптивних контрольно-навчальних програм;

– індивідуалізація навчання;

– використання методів адаптивного тестування як основи для реалізації методик для збільшення продуктивного навчання.

У технології адаптивного тестування можна виділити наступні напрями:

– розробка і створення теоретико-методологічного обґрунтування тестування, методів, алгоритмів і засобів технології адаптивного тестування;

– розробка адаптивного тестування як методу навчання і контролю, що є частиною загальної дидактичної системи навчання, що знижує частку викладацької праці і дозволяє викладачеві перерозподілити час навчання і оптимізувати навчальний процес;

– розробка на базі адаптивного тестування методології розвивального, продуктивного навчання і інших педагогічних інновацій;

– розробка методик, направлених на збільшення частки самостійного навчання в освітньому процесі;

– адаптивне тестування як методика для організації процесу самоосвіти;

– розробка програмно-інструментальних засобів, що забезпечують функціонування адаптивного тестування.

Оцінка якості навчання є одним з основних чинників підвищення ефективності освітнього процесу. В умовах забезпечення індивідуального підходу до кожного учасника тестування дуже важливо коректно провести якісний і кількісний аналіз його знань і умінь. Такий аналіз є складною багатофакторною залежністю з великим числом змінних. Проведення подібного аналізу часто вимагає великих витрат сил і часу на проведення статистичних розрахунків. Оцінювання якості навчання з використанням комп'ютерних технологій дозволяє значно скоротити час і витрати, підвищує інформативність результатів. Тому доцільно використовувати комп'ютерне адаптивне тестування (КАТ) [3].

2. Основні поняття та величини для адаптивного тестування

Однопараметрична модель Раша

Вводяться 2 основних параметри: θ – латентний параметр рівня знань учасника тестування і b – параметр складності завдання. Для однопараметричної моделі Раша використовується формула правильної відповіді на питання $P_j(\theta) = \dfrac{e^{(\theta - b_j)}}{1 + e^{(\theta - b_j)}}$.

Ймовірність правильної відповіді на питання залежить тільки від різниці між рівнем знань учасника тестування і складністю питання. Модель Раша не дозволяє питанням розрізнятися за дискримінацією: питання можна розташувати тільки за рівнем їх складності.

Двопараметрична модель Раша

Введемо початкові величини та поняття. Адаптивний тест складається із m питань. Нехай маємо змінні відповіді U_j , $j = 1,2,..,n$, які можуть приймати значення 1, якщо відповідь правильна, і 0, коли невірна, тобто

$$U_j = \begin{cases} 0, \text{якщо відповідь неправильна} \\ 1, \text{якщо отримана правильна відповідь} \end{cases}.$$

Логістична модель з двома параметрами (2PL модель):

$$P(U_j = 1 \mid \theta) = \frac{e^{a_j(\theta - b_j)}}{1 + e^{a_j(\theta - b_j)}},$$

де a_j – диференціююча здатність тестового j завдання (коефіцієнт дискримінації), b_j – складність завдання j ,θ – латентний параметр рівня знань учасника тестування. Будемо мати на увазі, що $i = 1,2..m; j = 1,2,..,n$, m – кількість учасників тестування, n – кількість тестових завдань.

Трипараметрична модель Бірнбаума

Згідно з трипараметричною логістичною моделлю (3PLM), ймовірність правильної відповіді $P(U_j = 1 \mid \theta)$ на j питання з рівнем знань θ знаходиться за формулою

$$P(U_j = 1 \mid \theta) \equiv c_j + (1 - c_j) \frac{e^{1.7 \cdot a_j(\theta_i - b_j)}}{1 + e^{1.7 \cdot a_j(\theta_i - b_j)}}, \qquad (1)$$

де c_j – ймовірність вгадування правильної відповіді при виконанні j завдання.

Різні вибори оцінки латентного параметра підготовленості θ

В адаптивному тестуванні можливо використати три етапи оцінки знань:

1 – оцінка знань, коли використовується процедура питання-відповідь;

2 – оцінка знань, коли протягом тесту використовувався вибір питань відповідно до параметра складності;

3 – оцінка знань оголошується в кінці тесту [4].

3. Реалізація адаптивного тестування на основі моделі Раша

За допомогою моделі Раша можна на основі тестувань визначити рівні складностей тестових завдань. Як наслідок, питання, які були задіяні в тесті, можна розмістити на одній шкалі. Але для шкалування всієї сукупності питань із бази тестів не можна просто застосувати модель Раша. Адже для одного тесту здійснюється лише вибірка окремої кількості питань. Необхідною умовою для застосування моделі Раша є те, що всі вибрані завдання повинні бути запропоновані кожному із студентів. В результаті ми отримаємо щільну матрицю результатів. Для шкалування наступної (відмінної від попередньої) вибірки питань на ту ж саму шкалу отримати логіти рівнів складності недостатньо, тому що при наступному тестуванні рівні знань студентів можуть відрізнятись і навіть склад групи дещо може різнитись від складу протестованих у попередньому тесті. Тому нами було запропоновано метод калібрування на основі трьох опорних завдань із попереднього тесту. Таким чином, у результаті скінченної кількості поточних тестувань можна звести на одну шкалу всі питання з бази. Кількість таких експериментальних тестувань залежить від вибірок.

Процес застосування даного методу зображено на схемі (рис. 1).

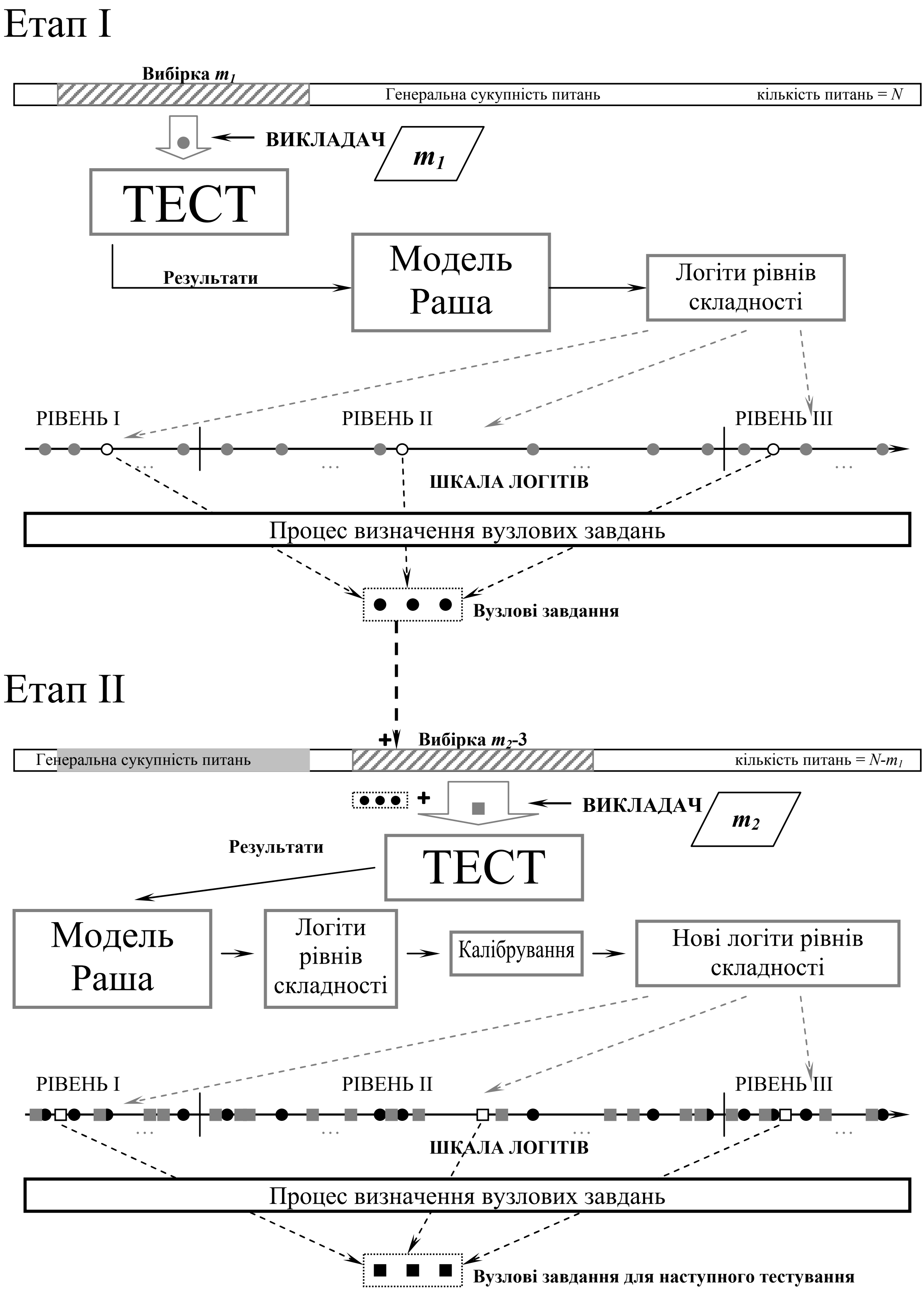


Рисунок 1 – Процес шкалування тестових завдань генеральної вибірки (I-II Етап)

Таким чином було сформульовано постановку задачі для практичної реалізації методу визначення міри складності тестового завдання на основі однопараметричної моделі Раша. Для реалізації цього методу нами було створено програмний модуль до існуючої системи дистанційного навчання за допомогою засобів PHP, який виконує контроль над подачею тестових питань, опрацьовує результати тестувань за допомогою методів, описаних вище, ставить у відповідність ідентифікатору питання значення логіта рівня складності. Це дозволяє розмістити питання на одній метричній шкалі.

Алгоритм функціонування такого модуля можна подати у вигляді блок-схеми, яка зображена на рис. 2.

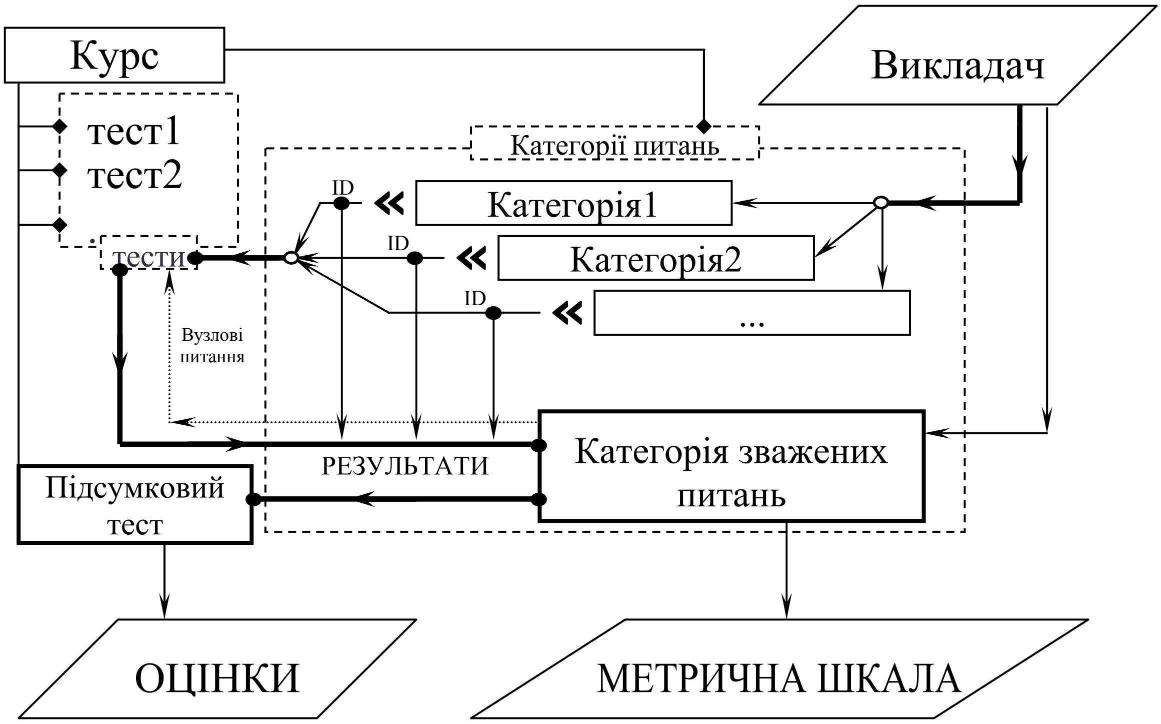


Рисунок 2 – Блок-схема програми

На перших стадіях функціонування дистанційного курсу формується перший тест для поточного тестування. Для тестового контролю в даному курсі викладач формує категорії завдань, із яких будуть вибиратися завдання для тестувань. Для першого тесту викладач задає кількість випадково відібраних завдань з цих категорій (рис. 3). На рисунку видно, що для тесту відібрано два завдання із категорії 1 і набір з десяти випадкових завдань з цієї ж категорії.

Після цього формується нова категорія «зважених» завдань, в котру ці завдання автоматично переносяться з наявних вже категорій. Відбувається організоване тестування. Після закінчення ми отримуємо щільну первинну матрицю результатів. Вона опрацьовується за допомогою моделі Раша. В результаті ми отримуємо масив логітів рівнів складності цих завдань.

При створенні наступних тестів викладач вибирає завдання із групи «незважених», що залишились в наявних категоріях. Таким чином, протягом поточних тестувань вичерпується вся генеральна сукупність тестових завдань, передбачених для даного курсу. Для того, щоб звести на одну шкалу логіти рівнів складності завдань, задіяних в другому тестуванні, перед тестуванням відбувається процес визначення вузлових завдань.

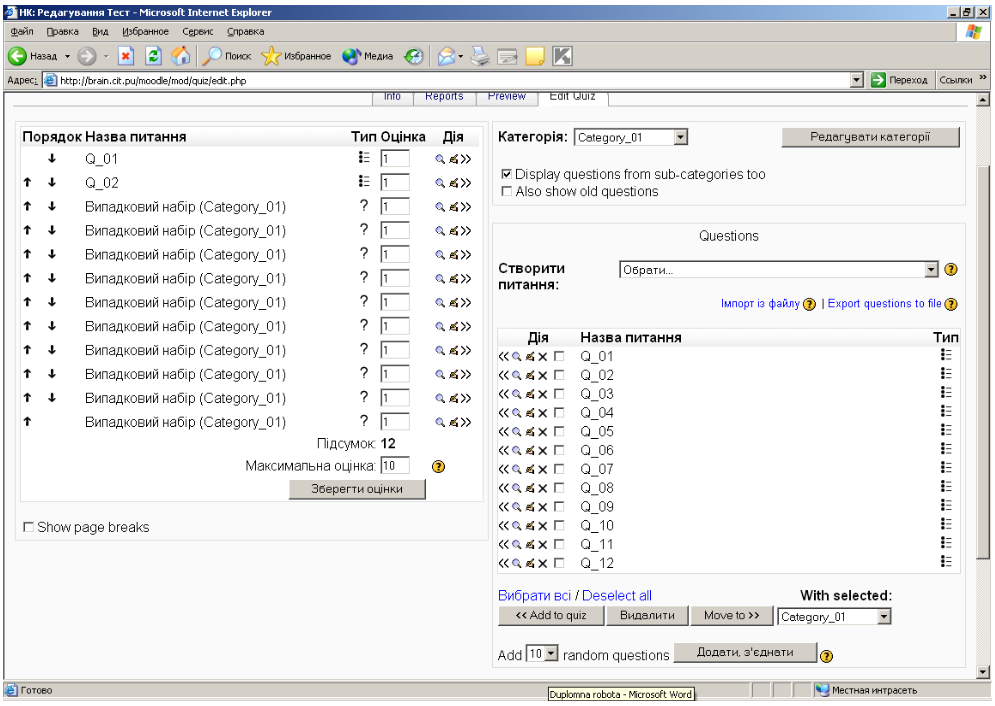


Рисунок 3 – Інтерфейс викладача

В даному експерименті це значення логітів, що набувають від’ємного, близького до нуля, додатного значень відповідно. Фактично, тепер ці завдання відповідають трьом рівням складності. За допомогою результатів відповідей на ці ж питання в другому тестуванні ми можемо визначити поправку для визначення логітів, які розмістяться на потрібній нам метричній шкалі. Процес вибору і переміщення завдань між категоріями аналогічний до попереднього. В експерименті нами було передбачено, що протягом поточних тестувань будуть зважені всі завдання з курсу і при підсумковому тестуванні можуть бути враховані ваги питань. Таким чином, ми вирішуємо проблему неадекватності визначення рівня знань студента. Маючи логіти, які відповідають завданням на метричній шкалі, такий метод надає можливість оцінити знання студента, враховуючи складність завдань, на які він відповідав.

Висновок

Розроблена модель адаптивного тестування дозволяє ефективно визначати рівень знань тих, хто тестується. Можна зробити узагальнення всіх переваг комп’ютерного адаптивного тестування:

1. КАТ уникає використання невідповідних питань. Питання, які занадто легкі або занадто важкі для учасників тесту, можуть викликати небажане поводження. Вони в значній мірі усунуті.

2. КАТ може бути коротшим від традиційного тестування.

3. КАТ може бути швидшим та інформативнішим. Неправильно вибране питання буде мати незначний вплив на тестові результати.

4. КАТ може бути чудовим засобом для правильного оцінювання учасника тесту і підбирання питань, відповідних до рівня знань.

5. КАТ дозволяє кожному учаснику тесту працювати індивідуально.

6. КАТ не потребує спеціального навчання учасників тесту.

7. КАТ пропонує багато варіантів для вибору індивідуального часу на відповідь конкретного питання.

8. КАТ зменшує в учасників тесту такі психологічні фактори, як втома і переживання.

9. КАТ може зменшити час тестування на більше ніж 50 % при підтримці того ж самого рівня надійності прийняття правильного рішення.

Література

1. Федорук П.І. Адаптивні тести: статистичні методи аналізу результатів тестового контролю знань // Математичні машини і системи. – 2007. – № 3,4. – С. 122-138.
2. Федорук П.І. Моделі і методи діагностики знань з використанням адаптивних тестів // УСиМ. – 2007. – № 5. – С. 68-76.
3. Ивлиев М.К. Разработка тестовых заданий для компьютерного тестирования: Учебно-методическое пособие. – М.: ИМПЭ им. А.С. Грибоедова, 2001. – 69 с.
4. Wim J. van der Linden, Cees A.W. Glas. Computerized Adaptive Testing: Theory and Practice. – Dordrecht, The Netherlands: Kluwer, 2000. – 323 p.

П.И. Федорук

Использование адаптивных тестов в интеллектуальных системах контроля знаний

В статье рассмотрено проблемы тестового контроля знаний в современных учебных системах. Методики и инструментарий классической теории тестов, несмотря на большие возможности данной технологии, не решают многих задач, которые ставит современный уровень развития образования. Адаптивные тесты в данном случае позволяют решать эти вопросы и могут быть эффективно использованы для решения любых задач оптимизации учебного процесса – оценки эффективности педагогических инноваций и технологий, мониторинга и т.д. Рассмотрено поход к конструированию тестов, который представлен в современной теории тестов на основе математической теории параметрической оценки тестових заданий на базе основных современных моделей: однопараметрической модели Раша, двупараметрической и трехпараметрической модели Бирнбаума.

P.I. Phedoruk

The Usage of Adaptive Tests in Intellectual Knowledge Control Systems

The article reveals the problems of test knowledge control in modern educational establishments. Methods and instruments of classic theory apart big possibilities do not solve many tasks, that gives modern level of education development. Adaptive tests in the given case let us to solve the given questions and may be used effectively for solving any tasks of learning process optimization- evaluation of effectiveness of pedagogical innovations and technologies, monitoring etc. The article describes the approach to test construction which is described in the modern test theory on the basis of mathematical theory of parametric evaluation of test tasks on the basis of modern mathematical models: oneparametric model of Rush, twoparametric and threeparametric model of Birnbaum.

Стаття надійшла до редакції 17.07.2008.